Selecting and Evaluating Conspicuous Character Patterns

Seiichi Uchida
Kyushu Univ., Japan

Ryoji Hattori
Kyushu Univ., Japan

Masakazu Iwamura
Osaka Pref. Univ., Japan

Shinichiro Omachi
Tohoku Univ., Japan

Koichi Kise
Osaka Pref. Univ., Japan

Abstract

Character detection in scenery images is a very difficult task. This paper describes a strategy of selecting character patterns for easier detection in scenery image. These character patterns, called conspicuous character patterns, are selected from character font sets according to a criterion that evaluates how the font has a larger distance from a non-character pattern distribution and, simultaneously, with a smaller distance from a general character pattern distribution. Qualitative and quantitative evaluations have been made to observe how the selected fonts are more conspicuous than other fonts.

1. Introduction

The first step of camera-based character recognition in scenery images is character detection (also called character localization). Generally speaking, any character detection scheme is defined by the following four factors:

(i) How do we determine whether the target area is a character area or not?
(ii) How do we assume the general property of character patterns?
(iii) How do we assume the general property of non-character patterns?
(iv) What features do we use for representing patterns?

For satisfactory performance of character detection, those factors should be defined appropriately.

Many researchers have proposed many detection schemes according to their own definition of the four factors [1, 2, 3, 4]. For example, in a scheme based on edge complexity, (i) determination is done by some degree of edge complexity, (ii) character patterns are assumed to have complex edges, (iii) non-character patterns are assumed to have less complex edges, and (iv) patterns are represented by edge features.

In spite of those past researches, character detection is a difficult open problem. At ICDAR2003, a competition on character detection was held [5] and the F-measure of the best method was reported as only 0.5. As shown by this contest result, detection performance is still not satisfactory.

The difficulty comes from the inevitable trade-off of the problem. If we want to detect all characters, we often have many false detections around complex edge regions (such as tree regions). Conversely, if we want to suppress those false detections, we may have many miss detections. It will be very difficult to relax this trade-off in the strategy of the past attempts.

This paper suggests a new research direction to escape the trade-off on this problem — preparation of special character patterns, called *conspicuous character patterns*, which can be easily detected. Ideally speaking, conspicuous character patterns are conspicuous enough to be detected without false and miss detections in any scenery image.

This paper describes a systematic method to prepare a conspicuous character pattern for a certain detection scheme. The proposed method *selects* the most conspicuous character pattern from various fonts. Automatic synthesis of conspicuous character patterns will be discussed elsewhere in near future.

Note that the research direction of preparing special character patterns for camera-based character recognition is not very new. Classic OCR/MICR fonts are specially designed to be recognized correctly by machines. In [6, 7], we can find updated versions of OCR fonts for camera-based character recognition. Two-dimensional bar codes, such as QR-code, are designed to be detected easily by cameras. Characters on signboards are often designed to be detected easily by humans.

The rest of this paper is organized as follows. In Section 2, the idea of conspicuous character patterns is outlined with naive examples. The close relation between conspicuous character patterns and the character detection scheme to be used is also discussed. Then in Section 3, the proposed method for selecting the most conspicuous font is detailed. In Section 4, experimental results are shown for evaluating the conspicuousness of the selected fonts qualitatively and quantitatively.

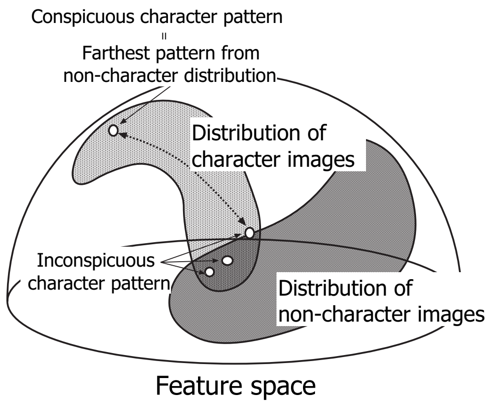

Figure 1. Character detection scheme assumed in this paper.

2. What Is Conspicuous Pattern?

2.1. Naive examples of conspicuous character pattern

A naive example of conspicuous character patterns is red-colored characters; we can detect them in a given scenery image only by seeking red areas in the image. Specifically, for detecting the red-colored characters, we will use a detection scheme where (i) a character area is determined if the color of the area is red enough, (ii) character patterns are assumed to be red-colored, (iii) non-character patterns are assumed not to be red-colored, and (iv) patterns are represented by RGB color feature. If we use this detection scheme, it is possible to detect all red-colored character patterns; that is, the red-colored characters are truly conspicuous for the detection scheme (even though there are false-alarms).

Another naive example is Times-Roman characters; we can detect them through a template matching-based search technique using original Times-Roman font templates. Similarly, any character font can be conspicuous if we employ a suitable detection scheme.

2.2. Relation between Character Detection Scheme and Conspicuous Character Pattern

These naive examples suggest an important fact that conspicuous character patterns should be designed together with the detection scheme. In other words, We must design the characters to be detected most easily for the detection scheme assumed. In fact, the red-colored characters are not conspicuous for the detector based on Times-Roman template matching. Similarly, the Times-Roman fonts are not conspicuous for the color-based detector.

3. Selecting Conspicuous Character Pattern

3.1. Detection Scheme

In this paper, we assume the following detection scheme as an example:

(i) A target area in a scenery image is determined as the character area when its distance to a character distribution is small and its distance to a non-character distribution is large.
(ii) The general property of characters is empirically defined by preparing a large set of font images. In other words, the property is defined as the actual distribution of font images.
(iii) The general property of non-characters is also empirically defined by preparing a large set of scenery image patches.
(iv) A normalized shape feature is used for representing patterns.

Figure 1 illustrates this detection scheme. Since any character or non-character pattern $\mathbf{p}$ is normalized so that $\|\mathbf{p}\| = 1$, the feature space becomes a hyper-sphere. Distributions of character patterns and non-character patterns are on the sphere. Then the target area is detected as a character area if it is far to the non-character distribution and near to the character distribution. The former condition is necessary to differentiate the character area from non-character patterns. The latter condition is introduced to guarantee human readability.

Note that this detection scheme is far more reasonable than the naive examples in Section 1. This is because the properties of characters and non-characters are defined empirically by using actual patterns. There is no strange assumption such that each character is red-colored.

3.2. Conspicuousness

Now, our task is to find the pattern which is the most easily detected by the assumed detection scheme of Section 3.1. Formally speaking, we must optimize the pattern $\mathbf{p}$ according to the criterion of maximizing its distance from the non-character distribution and, simultaneously, minimizing its distance from the character distribution.

According to the above criterion, we can define the *conspicuousness* of $\mathbf{p}$ by using its distances from the character and non-character distributions as follows. Let $\{\phi_1, \ldots, \phi_i, \ldots, \phi_I\}$ denote the principal I eigen-vectors of the covariance matrix of a distribution. Then the distance of a normalized pattern $\mathbf{p}$ from the distribution can be measured by using the I-dimensional subspace whose bases are

$\{\phi_1, \ldots, \phi_i, \ldots, \phi_I\}$. Specifically, we can have the distance of $\mathbf{p}$ as the canonical angle $\theta_{\mathbf{p}}$ given by

$$\cos^2 \theta_{\mathbf{p}} = \sum_{i=1}^{I} \frac{\langle \mathbf{p}, \phi_i \rangle^2}{||\mathbf{p}||^2 ||\phi_i||^2} = \sum_{i=1}^{I} \langle \mathbf{p}, \phi_i \rangle^2. \qquad (1)$$

Using the canonical angle, the conspicuousness of $\mathbf{p}$ is defined as

$$C_{\mathbf{p}} = \begin{cases} \theta_{\mathbf{p}}^N / (\theta_{\mathbf{p}}^N + \theta_{\mathbf{p}}^C) & \text{if } \theta_{\mathbf{p}}^N \geq \epsilon, \\ 0 & \text{otherwise,} \end{cases} \qquad (2)$$

where $\theta_{\mathbf{p}}^N$ is the canonical angle between $\mathbf{p}$ and the non-character subspace and $\theta_{\mathbf{p}}^C$ is the canonical angle between $\mathbf{p}$ and the character subspace. Both angles are derived by (1). The value $\theta_{\mathbf{p}}^N / (\theta_{\mathbf{p}}^N + \theta_{\mathbf{p}}^C)$ becomes large if $\theta_{\mathbf{p}}^N$ becomes large and/or $\theta_{\mathbf{p}}^C$ becomes small. The maximum and the minimum values of $C_{\mathbf{p}}$ are 1 (the most conspicuous) and 0 (the least conspicuous), respectively. The threshold ϵ (> 0) is introduced so that a small $\theta_{\mathbf{p}}^N$ does not give a large $C_{\mathbf{p}}$ with a very small $\theta_{\mathbf{p}}^C$.

The subspaces of the character and non-character distributions are determined by using empirical covariance matrices of many font images and many scenery image patches, respectively. Note that the dimensions I of the two subspaces can be different. Also note that it is possible to use any image feature (other than the bitmap feature) for representing those images, while we will use a shape feature, called direction code histogram [8], instead of the bitmap feature in order to minimize the effect of color and brightness. The detail of the shape feature will be given in Section 4.2.

3.3. Selection of most conspicuous character pattern

As noted in Section 1, we will not synthesize a conspicuous character pattern which globally maximizes $C_{\mathbf{p}}$. Instead, we will select the most conspicuous character pattern among various fonts. Specifically, we calculate the conspicuousness $C_{\mathbf{p}}$ for every font image $\mathbf{p}$, and then select the font image $\mathbf{p}$ with the largest $C_{\mathbf{p}}$ as the most conspicuous character pattern.

4. Experimental result

Using the above method with font images and scenery image patches, the selection and the evaluation of the most conspicuous character pattern were done experimentally. The following is the detail of this experiment.

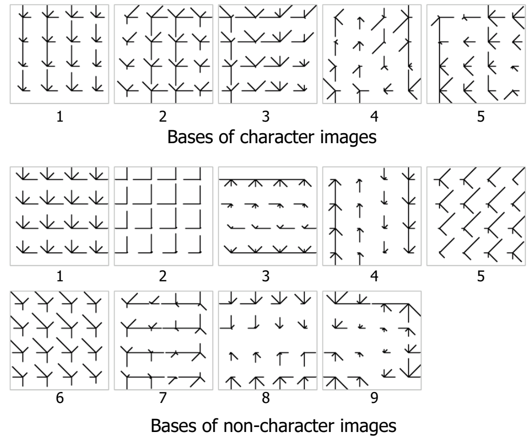

Figure 2. Bases of character distribution and non-character distribution.

4.1. Experimental setup

About 8,000 English capital letter images of 308 fonts were collected (26×308∼ 8,000). Those fonts were selected using their PANOSE. PANOSE is comprised of 10 elements and the first element represents the font family, such as “Latin Text”, “Latin Handwritten”, “Latin Decorative”, etc. The selected 308 fonts were of “Latin Text”, which is the font family suitable for general documents [1].

As non-character images, about 80,000 image patches (of 64×64 pixel size) were collected from scenery images of the CalTech’s background database[9], which is available as a public image database.

4.2. Feature space

Each image was represented as the direction code histogram [8]. First, each 64×64 image is converted into an edge map and divided into 16 blocks of 16×16 pixel size. The edge pixels in each block were quantized into four directions and then the number of edge pixels of each direction was counted within a block. Thus the dimension of the direction code histogram is 64 = 16(blocks)× 4(quantized directions). Since this feature is totally based on the edge direction, it is affected by neither color nor intensity.

Then a 64×64 covariance matrix was prepared for each of the character set and the non-character set and its I principal eigen-vectors were calculated as $\{\phi_1, \ldots, \phi_i, \ldots, \phi_I\}$. By observing the cumulative ratio,

[1]The authors found many “decorative” fonts with “Latin Text” PANOSE. The authors, however, did not remove them, because there is no objective criterion other than PANOSE for separating Latin Text fonts from decorative fonts.

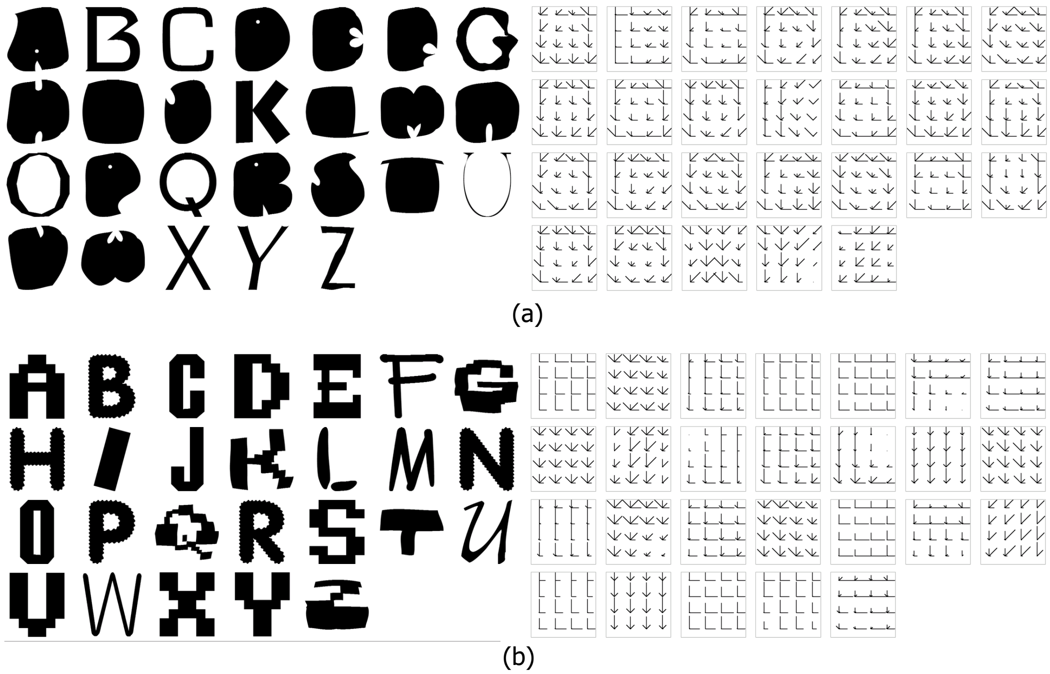

Figure 3. (a) Example of the most conspicuous pattern and (b) the least conspicuous pattern. Left: original font image. Right: direction code histogram.

Figure 4. Most and least conspicuous fontsets.

the number I was fixed at 5 for characters and 9 for non-characters. Figure 2 visualizes those eigen-vectors (i.e., the bases of the subspace) of characters and non-characters, respectively. Each eigen-vector is visualized by 64 short line segments. Each line segment represents the direction and the value of an element of the vector [2] .

[2] In Fig. 2, each element is visualized with its polarity; for example, if an element of $45°$ has positive (negative) value, it will be visualized as a $45°$ ($225° = 45° + 180°$) line segment.

4.3. Selected patterns

According to the procedure in Sections 3.2 and 3.3, the most conspicuous character pattern maximizing $C_{\mathbf{p}}$ was selected for each of 26 letter categories ("A" to "Z"). During the selection, the threshold ϵ was fixed at 0.4, which was determined experimentally.

Figure 3 (a) shows the most conspicuous character pattern of all categories. Round-shaped characters were selected as the conspicuous character patterns. This is because the direction code histogram of a round-shaped character has almost zero-value elements around the center area and thus is different from, especially, the bases ϕ_1 and ϕ_2 of non-character subspace. Consequently, $\theta^N_{\mathbf{p}}$ becomes large and thus $C_{\mathbf{p}}$ becomes large for the round-shaped characters.

It will be important to note that the conspicuous pattern is not limited to the pattern of Fig. 3 (a). Actually, the pattern of second-place was a similar round-shaped font. Therefore, we can use various round-shaped fonts as conspicuous patterns.

Figure 3 (b) shows the least conspicuous character pattern, which has the minimum $C_{\mathbf{p}}$ value by a small $\theta^N_{\mathbf{p}}$. Those characters often have a direction code histogram similar to one of the basis of non-character subspace. Thus $\cos^2 \theta^N_{\mathbf{p}}$ becomes large and, equivalently, $\theta^N_{\mathbf{p}}$ becomes small.

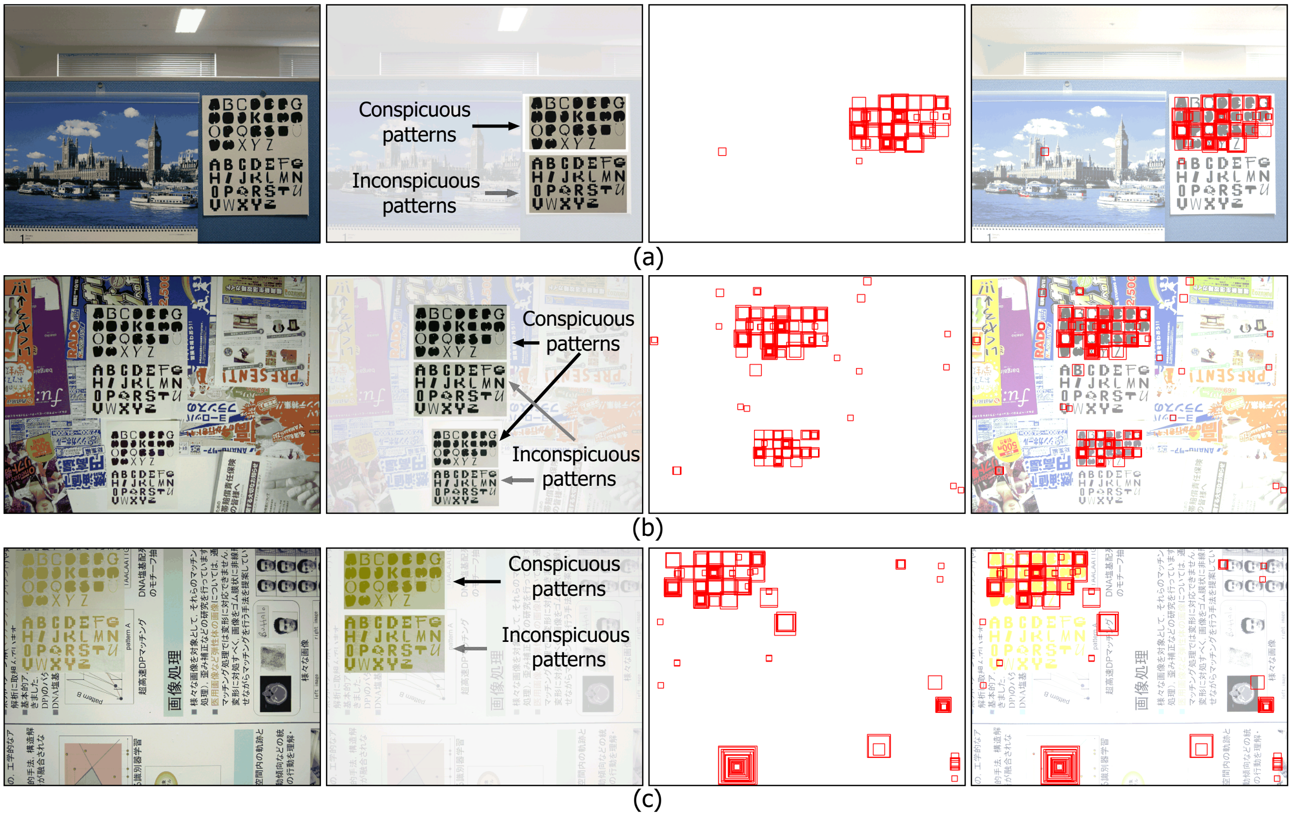

Figure 5. Character detection results on three different scenery images. From left to right: original scenery image with the most conspicuous and the least conspicuous character patterns (of Fig. 3), the locations of those patterns, detected areas, and the detected areas superimposed on the scenery image.

Figure 4 shows the most and the least "fontsets." That is, the sum of the conspicuousness $C_{\mathbf{p}}$ of all 26 categories from a fontset was used as criterion on selecting the most conspicuous fontset. (In Fig. 3, this selection was done at each individual letter.) While the most conspicuous fontset is very round and bold, the second most is rather natural and popular. It is interesting to note that even this natural fontset can be more conspicuous than other "gaudy" fontsets.

4.4. Qualitative evaluation

Actual easiness of detecting the selected conspicuous patters was examined. As shown in the leftmost figures in Fig. 5, the most conspicuous patterns were printed on a paper sheet and put in scene. Similarly, the least conspicuous patterns were also printed and put in the scene. The detection was performed by exhaustive search; the conspicuousness $C_{\mathbf{w}}$ was evaluated at all possible square windows, $\{\mathbf{w}\}$. Note that the size of square areas was also varied during the search. If $C_{\mathbf{w}}$ is larger than a threshold δ on a square window $\mathbf{w}$, we can consider that a conspicuous character pattern exists within $\mathbf{w}$. Note that for each result, the threshold δ was set at its optimal value in the sense that they give the maximum F-measures. (The definition of F-measure will be given later.)

Figure 5 (a) verifies that the most conspicuous patterns are really conspicuous; almost all the conspicuous patterns were detected successfully. (Only "B" was not detected here.) In addition, there are very few false alarms. This promising result indicates the importance of designing conspicuous patterns for the detection problem. Another important fact is that none of the least conspicuous patterns was detected. This fact is also promising because it proves the conspicuousness is very stable even in real scene.

Figures 5 (b) and (c) still verify the usefulness of the conspicuous patterns in very complex backgrounds. In (b), the conspicuous patterns were successfully detected regardless of the sizes. In (c), the conspicuous pattern were printed in different colors. Since the assumed feature (the direction code histogram) is affected by neither color nor intensity,

Table 1. Maximum F-measure on detecting conspicuous patterns.

criterion	$C_{\mathbf{p}}$	$\theta^N_{\mathbf{p}}$	$\theta^C_{\mathbf{p}}$
F-measure	0.73	0.39	0.60

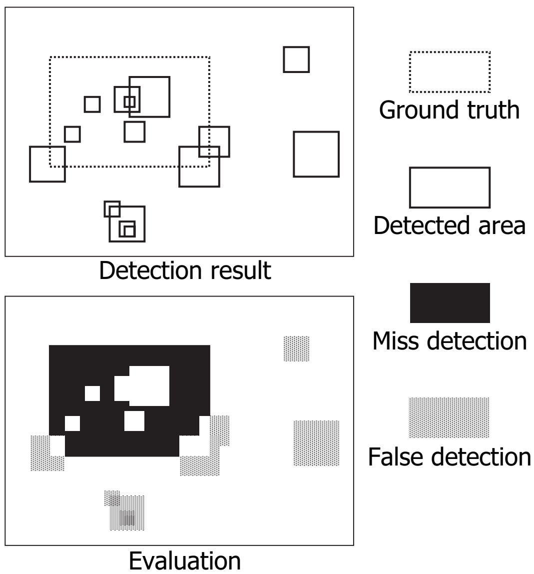

Figure 6. Quantitative evaluation method.

the colored conspicuous patterns were detected without obvious degradation. It is interesting to note that most characters in the backgrounds were not detected successfully. This also indicates that the characters designed to maximize their conspicuousness are very privileged patterns in the detection problem.

4.5. Quantitative evaluation

Quantitative evaluation of detection accuracy was done by the evaluation method shown in Fig. 6. Specifically, detection accuracy was evaluated by false detection rate $f = T/(R - U)$ and miss detection rate $m = S/U$, where

- R: the area of the entire image,
- U: the area of the ground truth (i.e., the area of the conspicuous character region) shown as a rectangular by dotted line,
- S: the area of the miss detection, shown as a black-filled region, and
- T: the area of the false detection, shown as a hatched region.

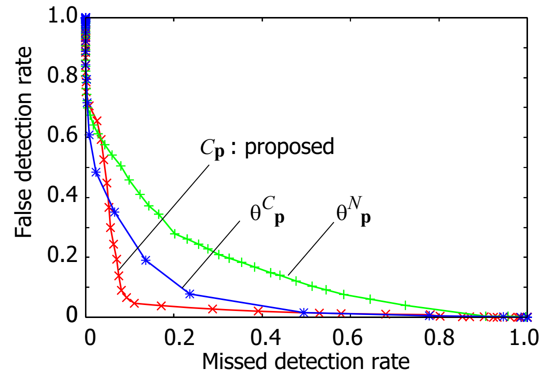

Figure 7. ROC curve on detecting conspicuous patterns.

Figure 8. Most conspicuous patterns selected under different criteria: (a) $\theta^N_{\mathbf{p}}$. (b) $\theta^C_{\mathbf{p}}$.

Figure 7 shows the ROC curve (i.e., the relation of f and m under different δ values) on the detection of the conspicuous character patterns selected by $C_{\mathbf{p}}$. The ROC curve becomes L-shaped one and thus indicates that the conspicuous character patterns can be detected accurately without false detection and miss detection.

Detection accuracy is also evaluated by the maximum of F-measure, which is defined as the harmonic mean of recall $r = (U - S)/U$ and precision $p = (U - S)/V$, i.e., $F = 2/(1/r + 1/p)$, where V is the area of the de-

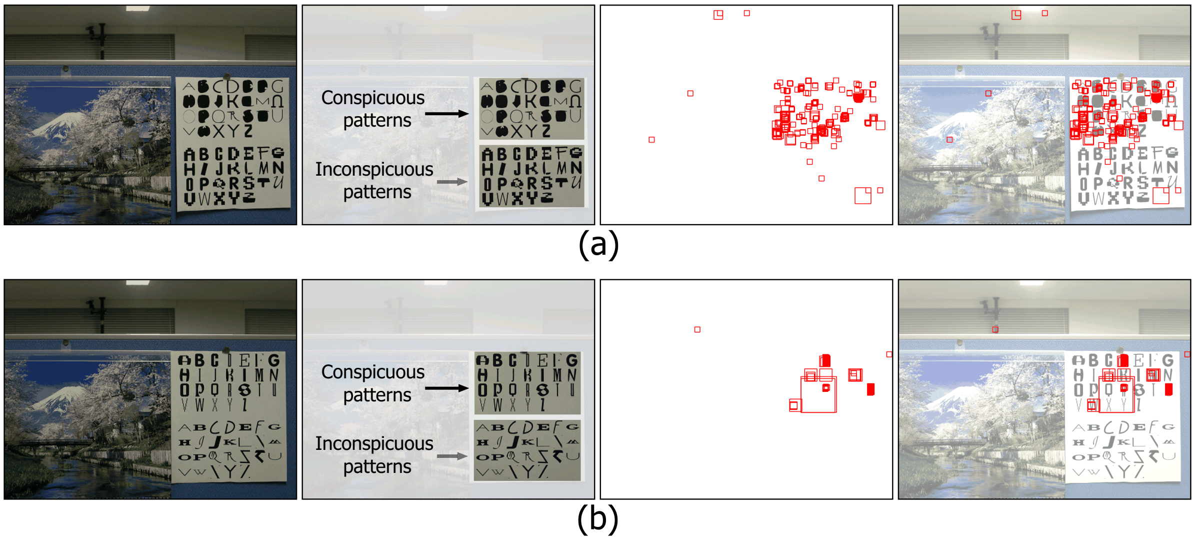

Figure 9. Detection results under different criteria: (a) $\theta^N_{\mathbf{p}}$. (b) $\theta^C_{\mathbf{p}}$.

tected region. The larger F-measure becomes, the better detection accuracy is. This is because a larger F-measure indicates that both of recall and precision could have better compromise. Table 1 shows that the maximum F-measure was 0.73 for $C_{\mathbf{p}}$. The goodness of this value will be understood through the comparative study in the next section.

4.6. Different criteria

Conspicuous patterns can be selected by using simpler criteria, such as $\theta^N_{\mathbf{p}}$ (farness from non-character pattern distribution) and $\theta^C_{\mathbf{p}}$ (closeness to character pattern distribution), instead of their combination $C_{\mathbf{p}}$. Figure 8 shows font patterns selected by using those criteria. The selected fonts are almost different from the fonts in Fig. 3, where $C_{\mathbf{p}}$ was used.

Figure 9 shows the detection results of those conspicuous patterns from scenery images. In those results, the threshold δ was fixed at its optimal value the sense that it gives the maximum F-measures. Comparing to Fig. 5, we can see far more miss detections in those results. Thus, the results indicate the usefulness of $C_{\mathbf{p}}$ over simple $\theta^N_{\mathbf{p}}$ and $\theta^C_{\mathbf{p}}$. This fact is confirmed by the ROC curves for $\theta^N_{\mathbf{p}}$ and $\theta^C_{\mathbf{p}}$ in Fig. 7 and F-measures in Table 1.

5. Conclusion

Like OCR/MICR fonts for easier character recognition, *conspicuous* character patterns were newly introduced for easier character detection in scenery images. This paper has discussed the approach of the conspicuous character pattern and detailed a systematic methodology of selecting the most conspicuous character font among various candidate fonts for a certain detection scheme. The actual conspicuousness of the selected pattern was confirmed qualitatively and quantitatively through several experiments; their detection was very easier than the detection of other characters.

This paper is the first attempt at developing conspicuous character patterns and therefore there are many future works. In this paper, we first assume a detection scheme and then design the conspicuous pattern for the scheme. In future, simultaneous design of the detection scheme and the conspicuous pattern should be investigated. Another future work is the systematic creation (synthesis) of the most conspicuous pattern instead of selection. For this future work, some parametric font-shape models will be necessary.

Acknowledgment

This work was supported in part by the Research Grant (No.20300049) of The Ministry of Education, Culture, Sports, Science and Technology in Japan.

References

[1] J. Keechul, K. I. Kwang, and A. K. Jain, “Text information extraction in images and video: a survey, Pattern Recognit., vol. 37, no. 5, pp. 977-997, 2004.

[2] R. W. Lienhart and F. Stuber, “Automatic text recognition in digital videos,” Proc. SPIE, vol. 2666 Image and Video Processing IV, pp. 180-188, 1996.

[3] X. Chen and A. L. Yuille, “Detecting and reading text in natural scenes,” Proc. CVPR, vol. 2, pp. 366-373, 2004.

[4] V. Wu, R. Manmatha, and E. M. Riseman, “An automatic system to detect and recognize text in images,” IEEE Trans. PAMI, vol. 21, no. 11, pp. 1224-1229, 1999.

[5] S. M. Lucas, et al., “ICDAR 2003 robust reading competitions: entries, results, and future directions,” IJDAR, vol. 7, pp. 105–112, 2005.

[6] S. Uchida, M. Sakai, M. Iwamura, S. Omachi, K. Kise “Extraction of embedded class information from universal character pattern,” Proc. ICDAR, vol. 1 of 2, pp.437-441, 2007.

[7] S. Omachi, M. Iwamura, S. Uchida, and K. Kise, “Affine invariant information embedment for accurate camera-based character recognition,” Proc. ICPR, vol. 2 of 4, pp. 1098-1101, 2006.

[8] F. Kimura, T. Wakabayashi, S. Tsuruoka, and Y. Miyake, “Improvement of handwritten Japanese character recognition using weighted direction code histogram,” Pattern Recognit., vol. 30, no. 8, pp. 1329-1337, 1997.

[9] `http://www.vision.caltech.edu/html-files/archive.html`